Random Latin squares

Peter J. Cameron

School of Mathematical Sciences
Queen Mary, University of London
Mile End Road
London E1 4NS, UK
p.j.cameron@qmul.ac.uk

RAND-APX meeting
Oxford, December 2003

Fisher and Yates

Fisher and Yates recommended that, in order to randomize an experimental design based on a Latin square, one should pick a random Latin square of the appropriate size.

Accordingly, they tabulated all Latin squares up to $n = 6$ (up to isotopy) and recommended choosing a random square from the tables and randomly permuting rows, columns and symbols.

Nowdays, this is no longer regarded as necessary for valid randomization. The row, column and symbol permutations suffice; any Latin square, however structured, will do.

On the other hand, we do now know how to choose a random Latin square . . .

A Latin square

A	I	Z	W	O	F	X	B	N	E	D	R	L	G	Q	U	C	K	M	V	Y	H	P	J	T	S
Y	L	R	U	T	H	D	Z	W	X	S	J	B	C	F	V	K	M	E	Q	G	I	N	P	O	A
X	O	B	Y	P	S	U	A	G	J	Z	E	C	F	H	D	N	I	K	W	Q	R	V	L	M	T
J	H	P	S	A	X	Y	K	L	Z	M	N	I	O	R	Q	V	D	F	T	B	C	W	G	U	E
G	B	E	Q	R	T	Z	F	H	Y	O	C	J	X	V	M	L	U	N	S	K	A	I	W	D	P
C	J	Q	F	K	O	H	V	U	D	T	G	R	A	Y	B	E	P	Z	L	N	X	S	M	W	I
N	K	D	O	F	U	P	S	A	B	W	V	G	Z	M	L	X	Q	T	E	C	J	Y	R	I	H
H	G	I	C	E	A	K	R	J	Q	L	O	N	S	B	W	Z	X	D	Y	F	V	M	T	P	U
W	M	S	A	D	Z	T	U	Q	R	X	B	P	E	O	F	G	Y	I	J	H	N	K	C	L	V
Q	E	K	L	G	B	M	W	S	P	C	U	Y	T	J	A	F	H	R	D	I	Z	O	N	V	X
P	U	Y	R	N	E	L	C	D	F	A	M	T	Q	G	I	H	J	V	O	Z	K	B	X	S	W
T	C	V	M	H	G	Q	D	O	N	U	X	E	R	W	P	B	A	L	I	S	F	J	K	Y	Z
M	T	N	Z	J	K	A	L	F	G	P	H	S	I	X	R	Y	W	U	C	V	E	D	O	Q	B
O	V	X	N	M	D	I	E	T	U	K	Q	W	Y	P	S	R	C	J	B	A	G	H	F	Z	L
L	S	T	H	I	C	W	Y	R	V	E	Z	D	J	K	X	U	N	P	G	M	Q	F	B	A	O
Z	R	A	E	B	V	S	X	K	I	Q	L	U	N	D	Y	W	G	O	F	P	T	C	H	J	M
B	X	C	K	L	Y	R	N	P	S	F	I	Z	H	T	O	M	V	W	U	E	D	Q	A	G	J
S	Y	H	I	X	W	J	O	B	M	G	D	V	K	Z	E	P	L	C	R	T	U	A	Q	N	F
E	D	F	V	Q	P	N	G	Z	A	B	W	O	U	I	J	T	R	Y	H	X	M	L	S	K	C
K	Z	G	X	Y	M	E	J	I	L	V	F	H	P	C	T	A	S	Q	N	O	W	U	D	B	R
R	Q	M	D	C	I	B	P	V	W	H	S	F	L	N	Z	J	T	X	A	U	O	G	Y	E	K
D	F	J	T	U	L	G	I	M	C	N	P	Q	V	A	K	O	B	H	Z	W	S	X	E	R	Y
U	P	O	B	Z	Q	V	H	C	K	R	Y	M	W	S	G	D	E	A	X	J	L	T	I	F	N
V	W	L	P	S	J	F	T	X	H	Y	A	K	D	E	N	I	O	G	M	R	B	Z	U	C	Q
F	A	U	J	W	N	O	M	E	T	I	K	X	B	L	C	Q	Z	S	P	D	Y	R	V	H	G
I	N	W	G	V	R	C	Q	Y	O	J	T	A	M	U	H	S	F	B	K	L	P	E	Z	X	D

The Jacobson–Matthews Markov chain

M. T. Jacobson and P. Matthews, Generating uniformly distributed random Latin squares, *J. Combinatorial Design* **4** (1996), 405–437.

Represent a Latin square as a function f from the set of ordered triples from $\{1,\ldots,n\}$ to $\{0,1\}$ such that, for any $x,y \in \{1,\ldots,n\}$, we have

$$\sum_z f(x,y,z) = 1,$$

with similar equations for the other two coordinates. Here $f(x,y,z) = 1$ means that the entry in row x and column y is z.

We allow also *improper Latin squares*, which are functions satisfying the displayed constraints but which take the value -1 exactly once (and the values 0 and 1 elsewhere). Now to take one step in the Markov chain starting at a function f, we do the following:

(a) If f is proper, choose (x,y,z) with $f(x,y,z) = 0$; if f is improper, start with the unique (x,y,z) such that $f(x,y,z) = -1$.

(b) Let x',y',z' be points such that

$$f(x',y,z) = f(x,y',z) = f(x,y,z') = 1.$$

(If f is proper, these points are unique; if f is improper, there are two choices for each of them.)

(c) Now increase the value of f by 1 on (x,y,z), (x,y',z'), (x',y,z'), and (x',y',z), and decrease it by 1 on (x',y,z), (x,y',z), (x,y,z'), and (x',y',z'). We obtain another proper or improper STS, according as $f(x',y',z') = 1$ or $f(x',y',z') = 0$ in the original.

Conjectures of Ryser and Brualdi

A *partial transversal* of a Latin square of order n is a set of cells, at most one in each row, at most one in each column, and at most one containing each symbol. It is a *transversal* if it has cardinality n (so that each "at most" becomes "exactly").

Ryser's Conjecture: If n is odd, then any Latin square of order n has a transversal.

Brualdi's conjecture: Any Latin square of order n has a partial transversal of cardinality $n-1$.

Of course random Latin squares are no help in proving these conjectures. But if they are false, one could look for random counterexamples. (Much better, though, to search in a more intelligent way: genetic algorithm??)

The theorem of Jacobson and Matthews asserts that the unique limiting distribution of this Markov chain is constant on proper Latin squares and on improper Latin squares.

So to choose a random Latin square, we start with any Latin square, repeat this procedure many times, and then continue until a proper Latin square is next obtained.

How many times? Bounds on the diameter of the graph are known, but little is known about the mixing time.

Two imprecise conjectures

- The number of rows of a random latin square of order n which are odd permutations is "approximately" Binomial$(n,\frac{1}{2})$;
- The second row of a normalised random latin square is "approximately" uniform on the derangements of $\{1,\ldots,n\}$.

In connection with the first conjecture, Häggkvist and Janssen showed that the probability that all rows are odd permutations is exponentially small (but not with the right constant).

The second conjecture can be generalised: perhaps, for k small relative to n, all $k \times n$ Latin rectangles are approximately equally likely as the first k rows of a Latin square.

The next two slides give the results of some experiments for $n = 10$.

Parity of rows

10240 random Latin squares of order 10.

# even rows	Observed	Expected
0	16	10
1	111	100
2	435	450
3	1206	1200
4	2162	2100
5	2512	2520
6	2112	2100
7	1146	1200
8	443	450
9	83	100
10	14	10

Random permutations

This detour gives the background for the derangement conjecture.

A *quasigroup* is an algebraic object whose Cayley table is an arbitrary Latin square.

Jonathan Smith has developed the character theory of quasigrouops (algebraic objects whose Cayley tables are arbitrary Latin squares). He observed that if the group generated by the rows and columns of the Cayley table is doubly transitive, then the character theory is “trivial”.

The following result shows that almost all quasigroups have trivial character theory.

Cycle structure of derangements

16481 random Latin squares of order 10.

Cycle structure	observed	expected
10	4433	4480
8,2	2793	2800
7,3	2152	$2133.\overline{3}$
6,4	1867	$1866.\overline{6}$
6,2,2	950	$933.\overline{3}$
5,5	914	896
5,3,2	1503	$1493.\overline{3}$
4,4,2	657	700
4,3,3	630	$622.\overline{2}$
4,2,2,2	245	$233.\overline{3}$
3,3,2,2	323	$311.\overline{1}$
2,2,2,2,2	14	$11.\overline{6}$

Łuczak and Pyber showed the following result:

Theorem Let g be a random permutation in S_n. The probability that there exists a transitive subgroup of S_n containing g, other than A_n and S_n, tends to zero as $n \to \infty$.

Corollary For almost all Latin squares, the group generated by the rows is the symmetric group S_n.

For the group generated by the rows is transitive and the first row is a random permutation; the result of Häggkvist and Janssen shows that the event “all rows even” has exponentially small probability, so the alternating group can be ignored.

Can we extend this result from quasigroups to loops (quasigroups with identity)? A loop is thus a structure whose Cayley table is a normalised Latin square (first row and column are the identity permutation).

It follows from the earlier results that for almost all derangements g (in the uniform measure), the only transitive group containing g is the symmetric or alternating group.

But for our purpose, we need the same assertion with the probability of g being the proportion of normalised random Latin squares having g as second row. If this is approximately uniform, the earlier argumemts could be used.

(a) If f is proper, choose xyz with $f(xyz) = 0$; if f is improper, start with the unique xyz such that $f(xyz) = -1$.

(b) Let x', y', z' be points such that

$$f(x'yz) = f(xy'z) = f(xyz') = 1.$$

(If f is proper, these points are unique; if f is improper, there are two choices for each of them.)

(c) Now increase the value of f by 1 on xyz, $xy'z'$, $x'yz'$, and $x'y'z$, and decrease it by 1 on $x'yz$, $xy'z$, xyz', and $x'y'z'$. We obtain another proper or improper STS, according as $f(x'y'z') = 1$ or $f(x'y'z') = 0$ in the original.

Problem: Is this chain connected? That is, can we move from any STS to any other by a sequence of such steps?

If so, then the limiting distribution is uniform on STSs. This is known to be true for $n \leq 15$.

Random Steiner triple systems

There is a Markov chain method for selecting a random STS from the uniform distribution – a small modification of the Jacobson–Matthews method for Latin squares.

Represent a STS as a function f from the set of unordered triples from $\{1,\ldots,n\}$ to $\{0,1\}$ such that, for any $x,y \in \{1,\ldots,n\}$, we have

$$\sum_{z \neq x,y} f(xyz) = 1.$$

We allow also *improper STSs*, which are functions satisfying the displayed constraint but which take the value -1 exactly once (and the values 0 and 1 elsewhere). Now to take one step in the Markov chain starting at a function f, we do the following: